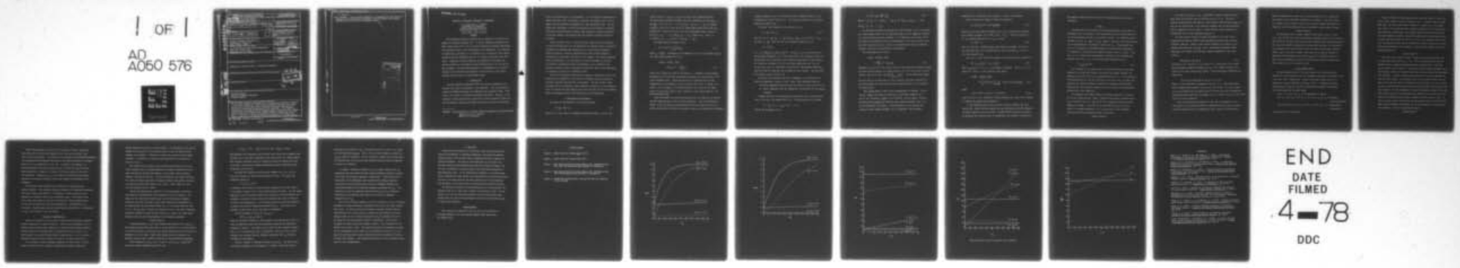
AD-A050 576 SOUTHERN METHODIST UNIV DALLAS TEX DEPT OF STATISTICS F/G 12/1
DELETION OF PRINCIPAL COMPONENTS IN REGRESSION.(U)
1978 R F GUNST, J T WEBSTER AFOSR-75-2871
UNCLASSIFIED AFOSR-TR-78-0152 NL

1 OF 1
AD
A050 576



END
DATE
FILMED
4-78
DDC

AD A050576

DDC FILE COPY

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

(19) REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER (18) AFOSR TR-78-0152	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) (6) DELETION OF PRINCIPAL COMPONENTS IN REGRESSION		5. TYPE OF REPORT & PERIOD COVERED (9) Interim rept.,
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) (10) R. F. Gunst [redacted] J. T. Webster		8. CONTRACT OR GRANT NUMBER(s) (15) AFOSR-75-2871
9. PERFORMING ORGANIZATION NAME AND ADDRESS Southern Methodist University Department of Statistics Dallas, Texas 75275		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 61102F (16) 2304/A5 (17) A5
11. CONTROLLING OFFICE NAME AND ADDRESS Air Force Office of Scientific Research/NM Bldg. 410, Bolling AFB Washington, D. C. 20332		12. REPORT DATE (11) 1978
		13. NUMBER OF PAGES (12) 24p.
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		15. SECURITY CLASS. (of this report) UNCLASSIFIED
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for Public Release: Distribution Unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

DDC RECEIVED MAR 1 1978 F

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Multicollinearity
Principal Component Regression
Biased Estimation

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Two techniques generally advocated for the deletion of principal components in regression analysis are: ~~(i)~~ delete components associated with small latent roots X'X, and ~~(ii)~~ delete components following nonrejection of a statistical test of the significance of the components. The estimator corresponding to the first procedure ~~(i)~~ is referred to as a restricted least squares estimator and that associated with the second ~~(ii)~~ is called a preliminary test estimator. Properties of these estimators are examined in this paper with special attention to the effects of multicollinearities on the preliminary

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

404 198. JOB

20. Abstract

est estimator. The restricted estimator is recommended for use unless inferences on the noncentrality parameter of the preliminary test clearly indicate that the test will have adequate power.

AFOSR-TR- 78-0152

DELETION OF PRINCIPAL COMPONENTS IN REGRESSION

R. F. Gunst and J. T. Webster
Department of Statistics
Southern Methodist University
Dallas, Texas 75275, U. S. A.

AFOSR-75-2871

SUMMARY

Two techniques generally advocated for the deletion of principal components in regression analysis are: (i) delete components associated with small latent roots of $X'X$, and (ii) delete components following nonrejection of a statistical test of the significance of the components. The estimator corresponding to procedure (i) is referred to as a restricted least squares estimator and that associated with (ii) is called a preliminary test estimator. Properties of these estimators are examined in this paper with special attention to the effects of multicollinearities on the preliminary test estimator. The restricted estimator is recommended for use unless inferences on the noncentrality parameter of the preliminary test clearly indicate that the test will have adequate power.

1. INTRODUCTION

Principal Component Regression (PCR) has long been employed in conjunction with tests of hypotheses of the components. This is advocated as a means of determining whether the components have predictive value. Massy (1965), Lott (1973), and others have proposed such tests. Recently, in the comparison of mean squared error properties of biased regression estimators, the PCR estimator has been used almost exclusively following a preliminary

---

KEYWORDS: Multicollinearity, Principal Component Regression, Biased Estimation

test of the predictivity of the components. In the simulation comparisons of Lawless and Wang (1976) and Dempster, Schatzoff, and Wermuth (1977), for example, the PCR estimators investigated all used preliminary tests. The uniformly poor performance of these PCR estimators led the authors to conclude that other biased estimators (notably ridge regression) should be preferred to the PCR estimator, particularly when the predictor variables are multicollinear.

Gunst and Mason (1977), on the other hand, report that a PCR estimator for which the predictivity of the components was assessed solely on the basis of whether the component was associated with a strong multicollinearity among the predictor variables outperformed the other biased estimators (including one ridge-type) with which it was compared. The authors suggested that the discrepancy in the performance of the PCR estimator in this and the previous investigations might be attributed to the instability of F statistics typically used in assessing the predictive merit of the components, an instability pointed out by Mansfield (1975).

This article addresses the question of whether a preliminary test of the predictivity of components in PCR should be attempted using the standard F statistics generally advocated for that purpose. The work of Bock, Yancy, and Judge (1973) on preliminary test estimation is central to this discussion since it provides the mean squared error (risk function) for the PCR estimator based on preliminary tests of the predictivity of the components.

2. PRELIMINARY TEST ESTIMATION

We employ the standardized linear regression model

$$\underline{Y} = \beta_o \underline{1} + X\underline{\beta} + \underline{\varepsilon}, \tag{2.1}$$

where $\underline{Y}$ is an (nxl) vector of observable response variables, $\underline{1}$ is an (nxl) vector of ones, $X = [\underline{X}_1, \underline{X}_2 \ldots, \underline{X}_p]$ is an (nxp) full-column-rank matrix of nonstochastic predictor variables that are standardized so that $\underline{1}'\underline{X}_j = 0$ and $\underline{X}_j'\underline{X}_j = 1$ for j=1, 2, ..., p, β_0 and $\underline{\beta}' = (\beta_1, \beta_2, \ldots, \beta_p)$ are unknown constants, and $\underline{\varepsilon}$ is an (nx1) vector of unobservable random error terms with $\varepsilon_i \sim NID(0,\sigma^2)$. Denote the latent roots and corresponding latent vectors of $X'X$ by $\ell_1 \geq \ell_2 \geq \ldots \geq \ell_p$ and $\underline{V}_1, \underline{V}_2, \ldots, \underline{V}_p$, respectively. Also, let $L = \text{diag}(\ell_1, \ell_2, \ldots, \ell_p)$ and $V = [\underline{V}_1, \underline{V}_2, \ldots, \underline{V}_p]$.

The least squares estimator of $\underline{\beta}$ is

$$\hat{\underline{\beta}} = (X'X)^{-1}X'\underline{Y} = \sum_{j=1}^{p} \ell_j^{-1} c_j \underline{V}_j, \tag{2.2}$$

where $c_j = \underline{V}_j' X' \underline{Y}$. One measure of the adequacy of (2.2) as an estimator of $\underline{\beta}$ is the (total) mean squared error of $\underline{\beta}$:

$$\begin{aligned} \text{mse}(\hat{\underline{\beta}}) &= E\{(\hat{\underline{\beta}}-\underline{\beta})'(\hat{\underline{\beta}}-\underline{\beta})\} \\ &= \sigma^2 \text{tr}(X'X)^{-1} = \sigma^2 \sum_{j=1}^{p} \ell_j^{-1}, \end{aligned} \tag{2.3}$$

where tr(A) denotes the trace of the matrix A. A drawback of least squares estimation of $\underline{\beta}$ is that when the predictor variables are multicollinear (2.3) can be extremely large. Multicollinearities are characterized as linear combinations of the columns of X that are nearly zero (see, e.g. Mason et al. (1975)). If the columns of X are multicollinear, one or more of the latent roots of X'X are very close to zero, resulting in one or more terms of (2.3) being extremely large.

Biased estimators of $\underline{\beta}$ have been developed with the intention of reducing the mean squared errors of the resulting estimators. Since the magnitude of $\text{mse}(\hat{\underline{\beta}})$ is for the most part controlled by the last few terms of (2.3) (i.e. by the latent roots of X'X that are closest to zero), one strategy for developing a biased estimator of $\underline{\beta}$ is to construct one that eliminates terms in (2.2) corresponding to small values of ℓ_j. The resulting estimator is a principal component estimator of $\underline{\beta}$.

To derive a PCR estimator of $\underline{\beta}$, rewrite (2.1) as

$$\underline{Y} = \beta_0\underline{1} + W\underline{\gamma} + \underline{\varepsilon}, \tag{2.4}$$

where $W = XV$, $W = [\underline{W}_1, \underline{W}_2, \ldots, \underline{W}_p]$ with $\underline{W}_j = X\underline{V}_j$, $\underline{\gamma} = V'\underline{\beta}$, and $\underline{\gamma}' = (\gamma_1, \gamma_2, \ldots, \gamma_p)$ with $\gamma_j = \underline{V}_j'\underline{\beta}$. Note that the least squares estimator of $\underline{\gamma}$ is

$$\hat{\underline{\gamma}} = L^{-1}W'\underline{Y};$$

i.e., $\hat{\gamma}_j = \ell_j^{-1}\underline{W}_j'\underline{Y}$ and $\hat{\gamma}_j \sim NID(\gamma_j, \ell_j^{-1}\sigma^2)$. The $\underline{W}_j$ in (2.4) are referred to as the principal components of X. PCR deletes some of the components from (2.4) and estimates the coefficients of the remaining components by least squares. Let $\hat{\underline{\gamma}}$ denote an estimator of $\underline{\gamma}$ for which some coefficients have been set to zero (note that this equivalent to deleting the corresponding components from (2.4)) and the remaining ones estimated by least squares. The associated PCR estimator of $\underline{\beta}$ is then $\hat{\underline{\beta}} = V\hat{\underline{\gamma}}$.

Basically two procedures have been recommended in the literature for selecting components to delete in PCR. Massy (1965) summarizes these as:

(i) delete components associated with the smallest latent roots of X'X, or

(ii) delete components that are unimportant as predictors of the response variable.

Suppose first that the components associated with the s smallest latent roots of X'X are to be deleted from (2.4). The PCR estimator of $\underline{\gamma}$ becomes

$$\hat{\underline{\gamma}}' = (\hat{\gamma}_1, \hat{\gamma}_2, \ldots, \hat{\gamma}_{p-s}, 0, 0, \ldots, 0),$$

and the PCR estimator of $\underline{\beta}$ is

$$\tilde{\underline{\beta}} = V\tilde{\underline{\gamma}} = V_L\tilde{\underline{\gamma}}_L = \sum_{j=1}^{p-s} \tilde{\gamma}_j \underline{V}_j, \tag{2.5}$$

where $V = [V_L:V_s]$, $V_L = [\underline{V}_1, \underline{V}_2, \ldots, \underline{V}_{p-s}]$, $V_s = [\underline{V}_{p-s+1}, \underline{V}_{p-s+2}, \ldots, \underline{V}_p]$,

and $\tilde{\underline{\gamma}}'_L = (\hat{\gamma}_1, \hat{\gamma}_2, \ldots, \hat{\gamma}_{p-s})$.

An equivalent procedure for deriving the PCR estimator (2.5) is to estimate $\underline{\beta}$ by least squares subject to the restriction $V'_s\underline{\beta} = \underline{0}$. Thus, this PCR estimator is a restricted least squares estimator. It is important to note that the restrictions are determined solely by an examination of $X'X$ and its latent roots and latent vectors and not as a result of inferences made using the response variable. Hence the restrictions are nonstochastic.

The mean squared error of $\tilde{\underline{\beta}}$ is

$$\text{mse}(\tilde{\underline{\beta}}) = E\{(\tilde{\underline{\beta}}-\underline{\beta})'(\tilde{\underline{\beta}}-\underline{\beta})\}$$

$$= \sigma^2 \sum_{j=1}^{p-s} \ell_j^{-1} + \underline{\beta}'V_sV'_s\underline{\beta}. \tag{2.6}$$

Comparison of (2.3) and (2.6) reveals that the restricted least squares estimator of $\underline{\beta}$ does indeed eliminate the largest terms of (2.3) but at the cost of introducing a term due to bias: $\underline{\beta}'V_sV'_s\underline{\beta} = \sum_{j=p-s+1}^{p} (\underline{V}'_j\underline{\beta})^2$. If the restrictions $V'_s\underline{\beta}=\underline{0}$ are true there is no bias term in (2.6); otherwise, the bias is nonzero and could potentially be larger than the terms eliminated from (2.3) by imposing the restrictions.

Next suppose Massy's (1965) second recommendation is adopted. In particular, suppose one wishes to delete the jth principal component if a test of the hypothesis $\gamma_j=\underline{V}'_j\underline{\beta}=0$ is not rejected. For the moment, consider testing jointly $\underline{\gamma}_s=V'_s\underline{\beta}=\underline{0}$ and using the least squares estimator of $\underline{\beta}$, (2.2), if this hypothesis is rejected. If the hypothesis is not rejected, the restricted least squares estimator (2.5) is employed. This estimator, referred to as a preliminary test estimator, is also a PCR estimator.

Often the hypothesis $V_s'\underline{\beta}=\underline{0}$ is tested by calculating

$$F_H = \hat{\underline{\beta}}'V_s(V_s'(X'X)^{-1}V_s)^{-1}V_s'\hat{\underline{\beta}}/sMSE, \qquad (2.7)$$

where $\hat{\underline{\beta}}$ is the least squares estimator of $\underline{\beta}$, (2.2), and MSE is the unbiased least squares estimator of σ^2, MSE $= \underline{Y}'\{I - n^{-1}\underline{1}\,\underline{1}' - X(X'X)^{-1}X'\}\underline{Y}/(n-p-1)$. If, for a preselected value of α, $F_H \geq c$, where

$$\alpha = \Pr\{F > c\} \qquad (2.8)$$

and F is a central F random variable with degrees of freedom s and (n-p-1), the hypothesis $V_s'\underline{\beta} = \underline{0}$ is rejected and $\hat{\underline{\beta}}$ is used to estimate $\underline{\beta}$. If $F_H < c$, $\tilde{\underline{\beta}}$ is the estimator of $\underline{\beta}$.

Bock et al. (1973) concisely represent the preliminary test estimator as

$$\bar{\underline{\beta}} = I_{[o,c)}(F_H)\tilde{\underline{\beta}} + I_{[c,\infty)}(F_H)\hat{\underline{\beta}}, \qquad (2.9)$$

where $I_{[a,b)}(\underline{u}) = 1$ if $a \leq u < b$ and equals 0 otherwise. From eq. (3.7) of Bock et al. (1973), in the notation of this paper,

$$\begin{aligned} mse(\bar{\underline{\beta}}) &= E\{(\bar{\underline{\beta}}-\underline{\beta})'(\bar{\underline{\beta}}-\underline{\beta})\} \\ &= \sigma^2 \sum_{j=1}^{p} \ell_j^{-1} - \sigma^2 P_1(\lambda) \sum_{j=p-s+1}^{p} \ell_j^{-1} + (2P_1(\lambda)-P_2(\lambda))\underline{\beta}'V_sV_s'\underline{\beta}, \qquad (2.10) \end{aligned}$$

where

$$P_j(\lambda) = \Pr\{F'(s,n-p-1,\lambda) < cs/(s+2j)\}, \qquad (2.11)$$

and $F'(s,n-p-1,\lambda)$ is a noncentral F random variable with s and (n-p-1) degrees of freedom and noncentrality parameter λ.

In the remaining sections of this paper we wish to compare the least squares $(\hat{\underline{\beta}})$, restricted least squares $(\tilde{\underline{\beta}})$, and preliminary test $(\bar{\underline{\beta}})$ estimators of $\underline{\beta}$ with specific attention focused on multicollinear predictor variables. By examining the characteristics of preliminary test estimator in particular, the apparent conflicts between the two PCR estimators maybe more clearly understood.

3. POWER

An appreciation of the effects of multicollinearities on the power of a preliminary test is important to the consideration of the relative merits of the three estimators defined in the previous section. For simplicity let us consider a test of H_o: $\underline{V}_p'\underline{\beta} = 0$ vs H_a: $\underline{V}_p'\underline{\beta} \neq 0$, where $\underline{V}_p$ is the latent vector of X'X corresponding to the smallest latent root, ℓ_p. In performing this test we are attempting to determine whether the pth principal component of X is important in predicting the response variable, since $\gamma_p = \underline{V}_p' \underline{\beta}$. If F_H (eg. (2.7)) is used as the test statistic, resulting in the uniformly most powerful test of H_o: $\gamma_p = 0$, the noncentrality parameter of F_H is

$$\lambda = \ell_p(\underline{V}_p' \underline{\beta})^2/2\sigma^2. \tag{3.1}$$

As a function of ℓ_p, observe that λ (and hence the power of the test) decreases as the multicollinearity indicated by $\underline{V}_p$ becomes stronger (for fixed $\underline{V}_p$, $\underline{\beta}$, and σ^2) since ℓ_p is approaching zero in (3.1). So the stronger the multicollinearity, the less the power of the test; yet F_H is often proposed for use when the predictor variables are extremely multicollinear. Qualitatively, this statistic appears to be a poor choice for assessing the predictivity of the components.

To illustrate the dramatic effects of multicollinearity on the power of the test quantitatively, Figure 1 exhibits power curves associated with F_H as a function of $\lambda/\ell_p = (\underline{V}_p' \underline{\beta})^2/2\sigma^2$ for a regression model with error degrees of freedom $\nu = n-p-1=10$ and two selections of ℓ_p:1.0 and 0.01. Since X is assumed standardized, ℓ_p=1 implies that X is an orthogonal matrix and no multicollinearities exist in the data.

[Insert Figure 1]

For fixed λ/ℓ_p (recall, $\lambda/\ell_p = (\underline{V}_p' \underline{\beta})^2/2\sigma^2$), Figure 1 reveals that the power drops precipitously when ℓ_p decreases from 1.0 to .01. Thus with strongly multicollinear data there is a much greater likelihood that $H_o:\underline{V}_p' \underline{\beta} = 0$ will not be rejected than if the predictor variables were not multicollinear. This is especially true for small α - levels and the results generalize to simultaneous tests of the hypotheses $H_o:V_s' \underline{\beta} = \underline{0}$.

In a more general setting than that considered in this paper, Toro - Vizcarrondo and Wallace (1968) studied preliminary testing from a slightly different viewpoint. Denote a least squares estimator of $\underline{\beta}$ subject to the general restrictions $K\underline{\beta} = \underline{m}$ by $\hat{\underline{\beta}}_R$. Toro - Vizcarrondo and Wallace (1968) define $\hat{\underline{\beta}}_R$ to be "better" than the least squares estimator $\hat{\underline{\beta}}$ if for every vector $\underline{d}$,

$$\text{mse}(\underline{d}' \hat{\underline{\beta}}_R) \leq \text{mse}\,(\underline{d}' \hat{\underline{\beta}}). \tag{3.2}$$

Translating this discussion to the problem being investigated in this paper, $\tilde{\underline{\beta}}$ (the principal component estimator with $\gamma_p = 0$) is "better" than $\hat{\underline{\beta}}$ if for every $\underline{d}$ (3.2) holds (replacing $\hat{\underline{\beta}}_R$ with $\tilde{\underline{\beta}}$). Toro-Vizcarrondo and Wallace also showed that

$$\text{mse}\,(\underline{d}' \tilde{\underline{\beta}}) \leq \text{mse}(\underline{d}' \hat{\underline{\beta}}) \text{ for all } \underline{d} \iff \lambda \leq 1/2,$$

where λ is the noncentrality parameter of F_H (eg. (2.7). They then proved that a uniformly most powerful test of $H_o:\lambda \leq 1/2$ vs $Ha:\lambda > 1/2$ is to reject H_o if F_H is greater than the upper $100(1-\alpha)\%$ critical point of a noncentral F distribution with noncentrality parameter 1/2, i.e. reject H_o if $F_H > F_\alpha'(1,\nu;\lambda=\frac{1}{2})$ and do not reject otherwise.

Since the noncentrality parameter of this test is identical to (3.1), the effects of multicollinear predictor variables will be the same as those of the more traditional test of comparing F_H with a central F critical point.

Figure 2 displays power curves for the test proposed by Toro - Vizcarrondo and Wallace for the same model parameters as in Figure 1. The power curves are slightly lower in Figure 2 than those in Figure 1 and the debilitating influence of strong multicollinearities on the power is again clearly evident.

[Insert Figure 2]

One conclusion that is readily apparent from Figures 1 and 2 is that $\lambda/\ell_p = (\underline{V}_p' \underline{\beta})^2/2\sigma^2$ must be very large before reasonable power will be achieved with either of the above tests when the data contains strong multicollinearities. Thus when the predictor variables are strongly multicollinear, the preliminary test estimator will tend to reduce to the restricted least squares estimator unless λ/ℓ_p is very large; however, even when λ/ℓ_p is small or moderate in magnitude, the preliminary test rejects $H_0:\underline{V}_p' \underline{\beta} = 0$ frequently enough to make it more advantageous to use the restricted least squares estimator for these values of λ/ℓ_p, as we shall now see.

4. MEAN SQUARED ERRORS

The mean squared errors of the least squares (LS), restricted least squares (R), and preliminary test (PT) estimators were given in equations (2.3), (2.6), and (2.10), respectively. Graphs of the mean squared errors as a function of λ/ℓ_p are exhibited in Figures 3 and 4. In computing the mean squared errors, three p=5, n-p-1=10 X matrices were studied, each defined by the following sets of latent roots:

(i) $\ell_1 = \ell_2 = \ell_3 = \ell_4 = \ell_5 = 1.00$ (orthogonal X matrix)

(ii) $\ell_1 = 2.90$, $\ell_2 = 1.00$, $\ell_3 = 0.70$, $\ell_4 = 0.30$, $\ell_5 = 0.10$ (moderate multicollinearity)

(iii) $\ell_1 = 2.99$, $\ell_2 = 1.00$, $\ell_3 = 0.70$, $\ell_4 = 0.30$, $\ell_5 = 0.01$ (strong multicollinearity).

In all cases $\sigma^2 = 1.0$ was used.

Figure 3 displays the mean squared error curves for models (i) and (iii) over the same range of λ/ℓ_p as in Figures 1 and 2. Consider first the orthogonal data, the lower three curves in Figure 3. Unless λ/ℓ_p is very small (less than 0.5 for restricted least squares and about 0.4 for the preliminary test estimator), least squares has a smaller mean squared error than either principal component estimator, although the difference in mean squared errors is never very great between LS and PT. If λ/ℓ_p is not extremely small, moreover, the restricted least squares estimator is clearly inferior to the least squares and preliminary test estimators. But our major concern here is with multicollinear data.

[Insert Figure 3]

The upper three curves of Figure 3 display the mean squared errors for the strong multicollinearity ($\ell_p = .01$). The relationships among the mean squared errors is completely changed from the orthogonal data. Over the entire range of λ/ℓ_p in Figure 3, the restricted least squares estimator has a substantially smaller mean squared error than least squares and the preliminary test estimator. Further, the mean squared error for PT is much smaller than that for least squares. From Figures 1 and 2 one can again observe that the power of the test of either $\lambda=0$ (Figure 1) or $\lambda\leq1/2$ (Figure 2) is low, accounting for the smaller mean squared error for the preliminary test estimator than least squares. Yet the hypotheses are rejected frequently enough to force the mean squared error for the preliminary test estimator to be much larger than that of the restricted least squares estimator over this range of λ/ℓ_p. In fact λ/ℓ_p must be quite large before the mean squared error of the restricted least squares estimator will exceed that of least squares or the preliminary test estimator. This is illustrated in Figure 4.

[Insert Figure 4]

Model configurations (ii) and (iii) are plotted in Figure 4 (note that both curves for the restricted estimator are so close to one another that only one has been graphed). The effects of the strength of the multicollinearities can be appreciated by comparing the curves as the multicollinearity is strengthened (e.g. as ℓ_p changes form .10 to .01). In general, the stronger the multicollinearity the wider the range of λ/ℓ_p for which the restricted least squares estimator is superior (in a mean squared error sense) to the other two estimators. Conversely, if ℓ_p is not small the restricted least squares estimator can be greatly inferior to the least squares and preliminary test estimators.

With strongly multicollinear data, therefore, the restricted least squares estimator - the principal component estimator with components associated with small latent roots deleted - is preferable to least squares or a preliminary test estimator unless λ/ℓ_p is extremely large. From Figures 1 and 2 note that large values of λ/ℓ_p are also required to insure adequate power for the test of predictivity of multicollinearities. Unfortunately, λ/ℓ_p is unknown. Whether sample data can provide useful information on the magnitude of λ/ℓ_p is the subject of the next section.

5. PRACTICAL CONSIDERATIONS

Ideally, we should not use the data to indicate which regression estimator should be employed in a specific analysis. Since using a preliminary test to decide whether to employ least squares or a restricted least squares estimator actually produces an estimator that is a mixture of the two, so too if one must estimate λ/ℓ_j before deciding whether to employ LS, R, or PT, the actual estimator utilized is also a mixture of LS and R. So what is one to do?

In the absence of exact knowledge regarding the value of λ/ℓ_j, we still prefer to allow the data to suggest an appropriate estimator despite the mixture problems, but only to a limited extent. If inferences on λ/ℓ_p clearly suggest that the power of the preliminary test is good, we employ the preliminary test estimator. Otherwise, we prefer the restricted least squares estimator. An example will illustrate our application of the results of Sections 3 and 4.

The pitprop data of Jeffers (1967) concerns the construction of a prediction equation for the maximum compressive strength of timbers from information provided by thirteen measurements on each timber (such as length, diameter, and specific gravity). The data collected on 180 timbers indicates that there are three strong multicollinearities among the predictor variables as evidenced by three small latent roots of $X'X$: .0387, .0414, and .0506 (the next smallest root is .1908).

Scaled mean squared errors (mse/σ^2) of the least squares, restricted least squares, and preliminary test estimators are given in Figure 5. For simplicity the scaled mean squared errors of the two principal component estimators are drawn as though a single small latent root of magnitude .05 is being deleted (all three small latent roots of the pitprop data are close to .05). We are going to examine the deletion of each of the three components separately although we could consider them two at a time or all three simultaneously with only minor modifications of the following arguments.

[Insert Figure 5]

As mentioned earlier, F_H is the uniformly most powerful test of $\underline{V}_j' \underline{\beta} = 0$. The previous sections have shown that it is also effective in selecting whether to use least squares or restricted least squares provided that the noncentrality parameter is not too small. Rather than blindly using the statistic, we advocate using the data to obtain information on the noncentrality parameter.

Point estimates of λ_j/ℓ_j, j=11, 12, and 13, using $\lambda_j/\ell_j = (\underline{V}_j' \hat{\underline{\beta}})^2/2\hat{\sigma}^2$ with least squares estimates of $\underline{\beta}$ and σ^2, are

$$\hat{\lambda}_{11}/\ell_{11} = 4.13, \quad \hat{\lambda}_{12}/\ell_{12} = 335.21, \text{ and } \quad \hat{\lambda}_{13}/\ell_{13} = 0.001.$$

Unfortunately, the estimators used to obtain these values have extremely large variances due to the small magnitudes of the latent roots (cf. Silvey (1969)). Thus the point estimates cannot be trusted and interval estimates will tend to be wide. The interval estimates nevertheless provide interesting implications for this data set.

Two-sided 95% confidence intervals were computed for λ_j/ℓ_j, j=11-13, using the noncentral F distribution associated with F_H. For λ_{11}/ℓ_{11} the confidence interval is

$$0 \leq \lambda_{11}/\ell_{11} \leq 66.67 .$$

As expected, this interval is quite wide and includes not only the range of λ/ℓ included in Figure 5 but it also extends beyond that range. For the reasons expressed in the two previous sections for being wary of the preliminary test estimator, we prefer to use the restricted estimator when there is doubt as to which is more appropriate. So we advocate using the restricted estimator to delete the principal component associated with ℓ_{11}. The choice of an estimator is more clear cut for the other two components.

The 95% confidence interval for λ_{12}/ℓ_{12} is

$$125.41 \leq \lambda_{12}/\ell_{12} \leq 643.53.$$

Again the confidence interval is wide; however, the lower bound is about an order of magnitude larger than the preferential range for the restricted estimator in Figure 5. The power of F_H is good over this interval (greater than .8) so the preliminary test is recommended. With F_H =27.80, the preliminary test indicates that the component associated with ℓ_{12} should be retained in the estimator.

Finally, consider a confidence interval for λ_{13}/ℓ_{13}. The actual value of the point estimate of this parameter is 0.000986. Using the central F distribution associated with F_H, the observed value of F_H falls in the lower 1% of the population <u>for $\lambda=0$</u>. Thus a 95% two-sided confidence interval for λ_{13}/ℓ_{13} cannot be calculated. We can, therefore, conclude with a high degree of certitude that the restricted least squares estimator should be employed to delete this component.

In summary, confidence intervals on λ_j/ℓ_j clearly indicate that the preliminary test and restricted least squares estimators should be utilized on the last two principal components, respectively. Concern over the reduced power of the preliminary test estimator when testing components associated with multicollinearities lead us to recommend employing the restricted estimator on the principal component associated with ℓ_{11}. The conclusion is that the components associated with ℓ_{11} and ℓ_{13} should be deleted but the preliminary test forces us to retain the component corresponding to ℓ_{12}.

This use of confidence intervals to obtain information on λ/ℓ_j is actually equivalent to simply performing a preliminary test with a very small significance level. By insisting that the confidence intervals provide relative certitude that λ/ℓ_j is large enough to enable the preliminary test to have adequate power, we are in effect insisting that F_H be extremely far out in the upper tail of the associated central F distribution. What we have attempted to convey through this example, however, is our rationale for demanding such a small α level. The previous sections have presented the basis for our recommendation that unless one is relatively sure that λ/ℓ_j is not small the restricted least squares estimator should be preferred to the preliminary test estimator. The simulations referred to in the Introduction also attest to this recommendation.

6. CONCLUSION

Motivated by discrepancies in conclusions drawn from simulations comparing the performance of regression estimators, this paper has examined characteristics of the two most widely recommended principal component regression estimators. The power of the preliminary test was shown to be severely reduced by multicollinearities among the predictor variables, yet the test is proposed to ascertain whether these same multicollinearities have predictive value. If the noncentrality parameter is not sufficiently large to dominate the small latent root associated with the multicollinearity, not only is the power poor but the mean squared error of the preliminary test estimator is also larger than the restricted least squares estimator.

The deliterious effects of multicollinearities on the preliminary test estimator lead one to infer that it should not routinely be used. Our preference is to use the restricted estimator unless the data yields clear evidence that the power of the preliminary test will be sufficiently large to render good inferences.

ACKNOWLEDGEMENT

This research was supported in part by the Air Force Office of Scientific Research, Air Force Systems Command, USAF, under grant No. AFOSR-75-2871.

FIGURE CAPTIONS

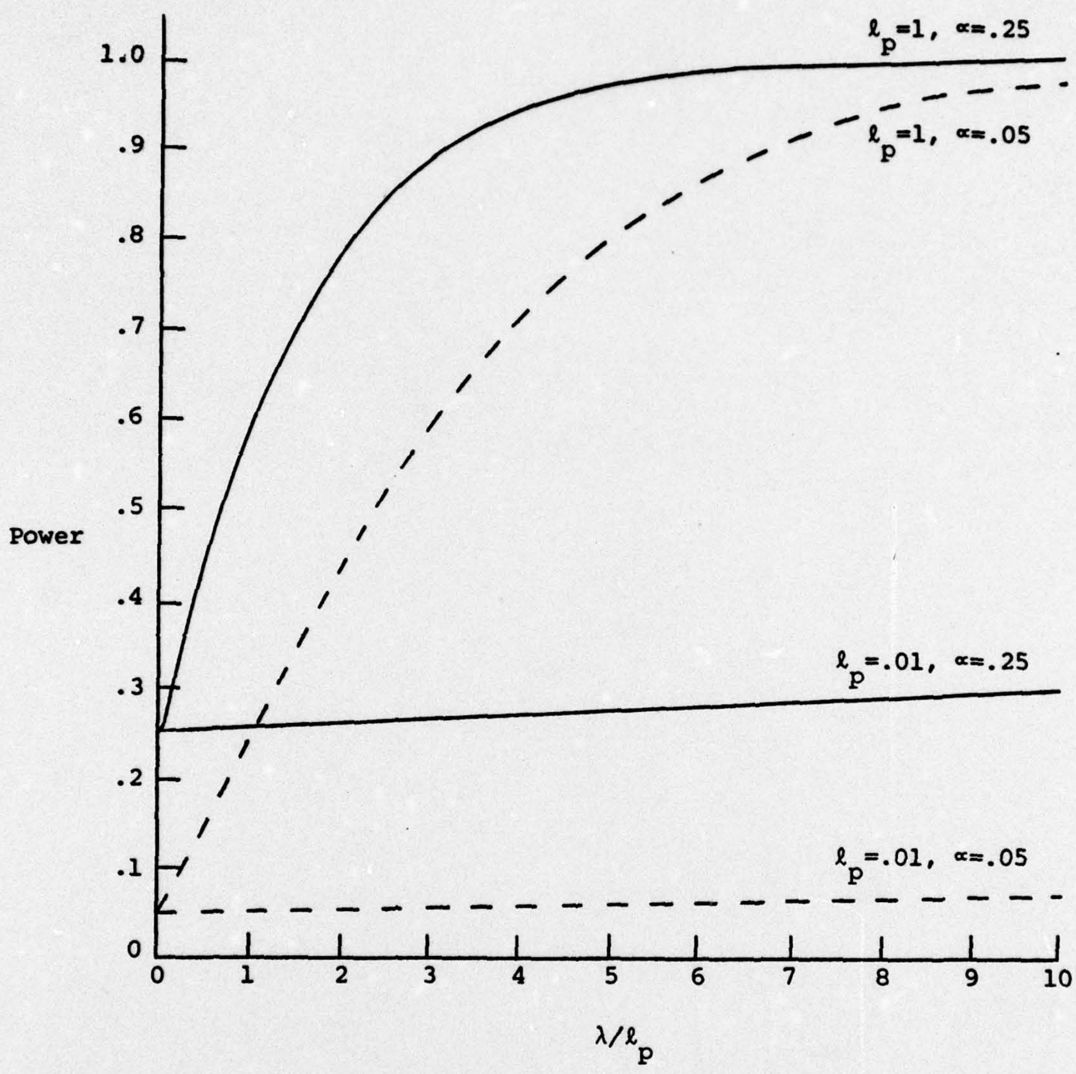
Figure 1. Power Curves For Testing $\underline{V}_p'\beta=0$ (ν=10).

Figure 2. Power Curves For Testing $\lambda\leq\frac{1}{2}$ (ν=10).

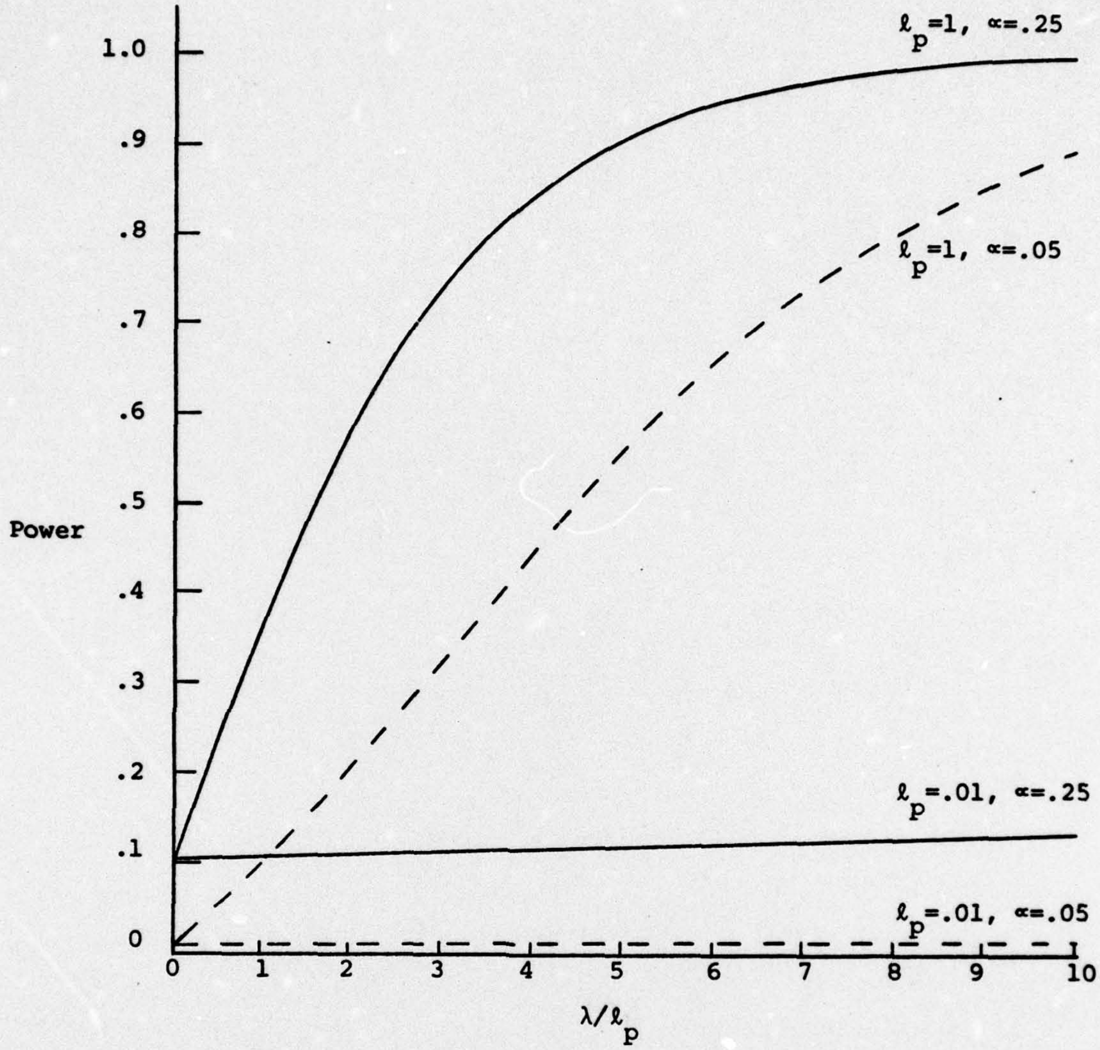
Figure 3. Mean Squared Errors For Least Squares (LS), Preliminary Test Test (PT), And Restricted Least Squares (R) Estimators.

Figure 4. Mean Squared Errors For Least Squares (LS), Preliminary Test (PT), And Restricted Least Squares* (R) Estimators.

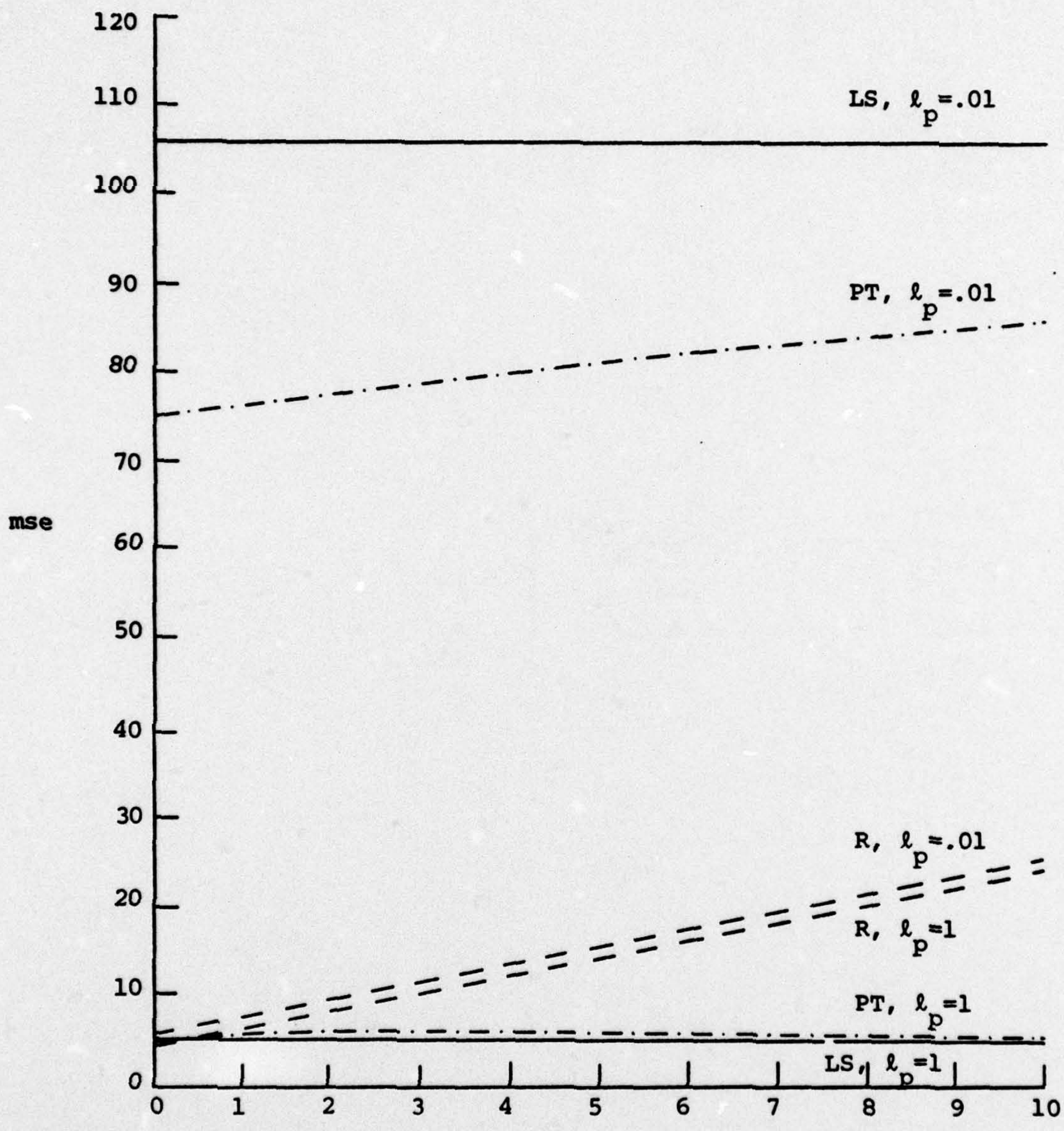
Figure 5. Scaled Mean Squared Errors, Pitprop Data With One Component (ℓ=.05) Deleted.



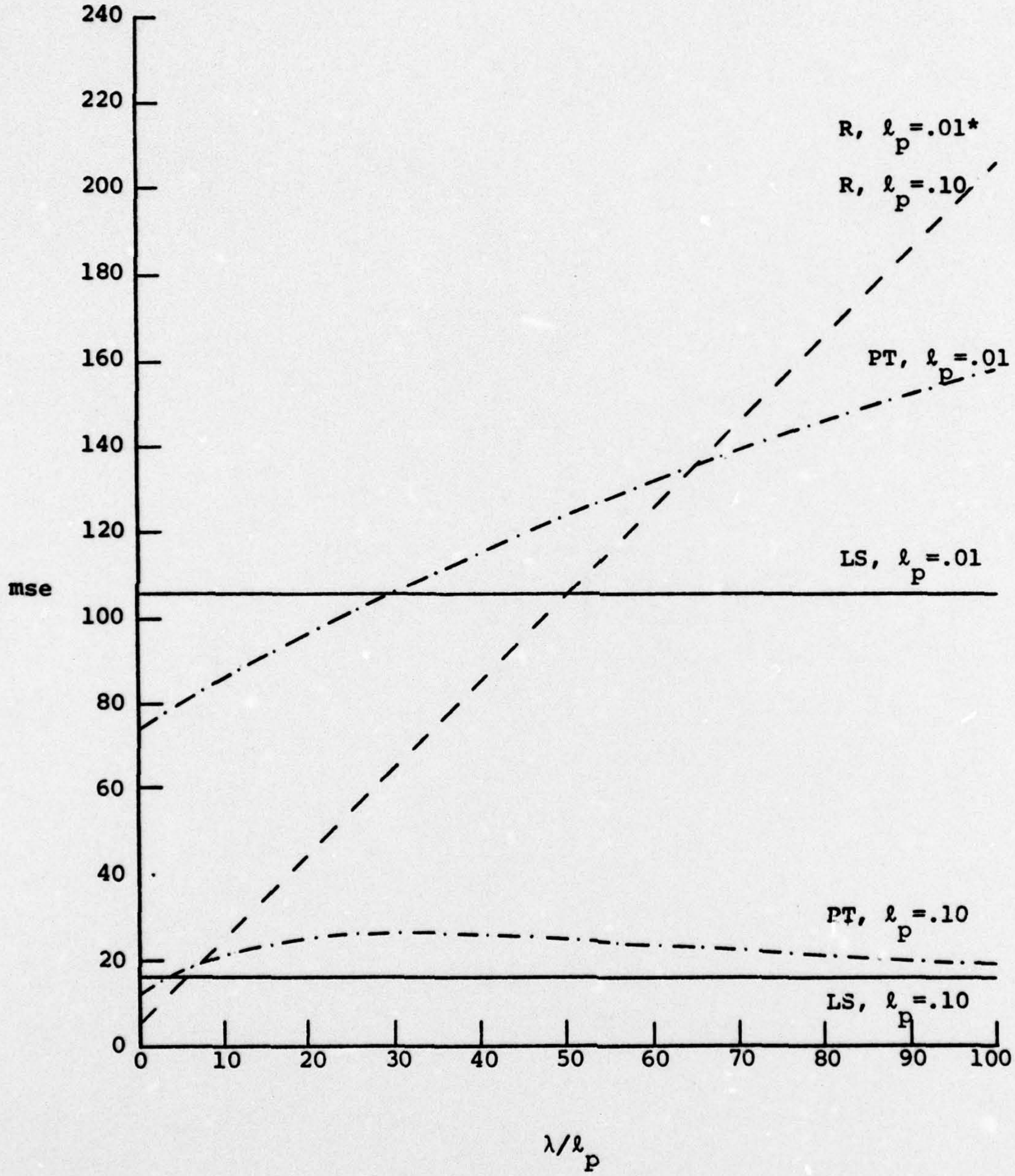
ℓ_p=1, α=.25
ℓ_p=1, α=.05
ℓ_p=.01, α=.25
ℓ_p=.01, α=.05
Power
1.0
.9
.8
.7
.6
.5
.4
.3
.2
.1
0
0 1 2 3 4 5 6 7 8 9 10
λ/ℓ_p



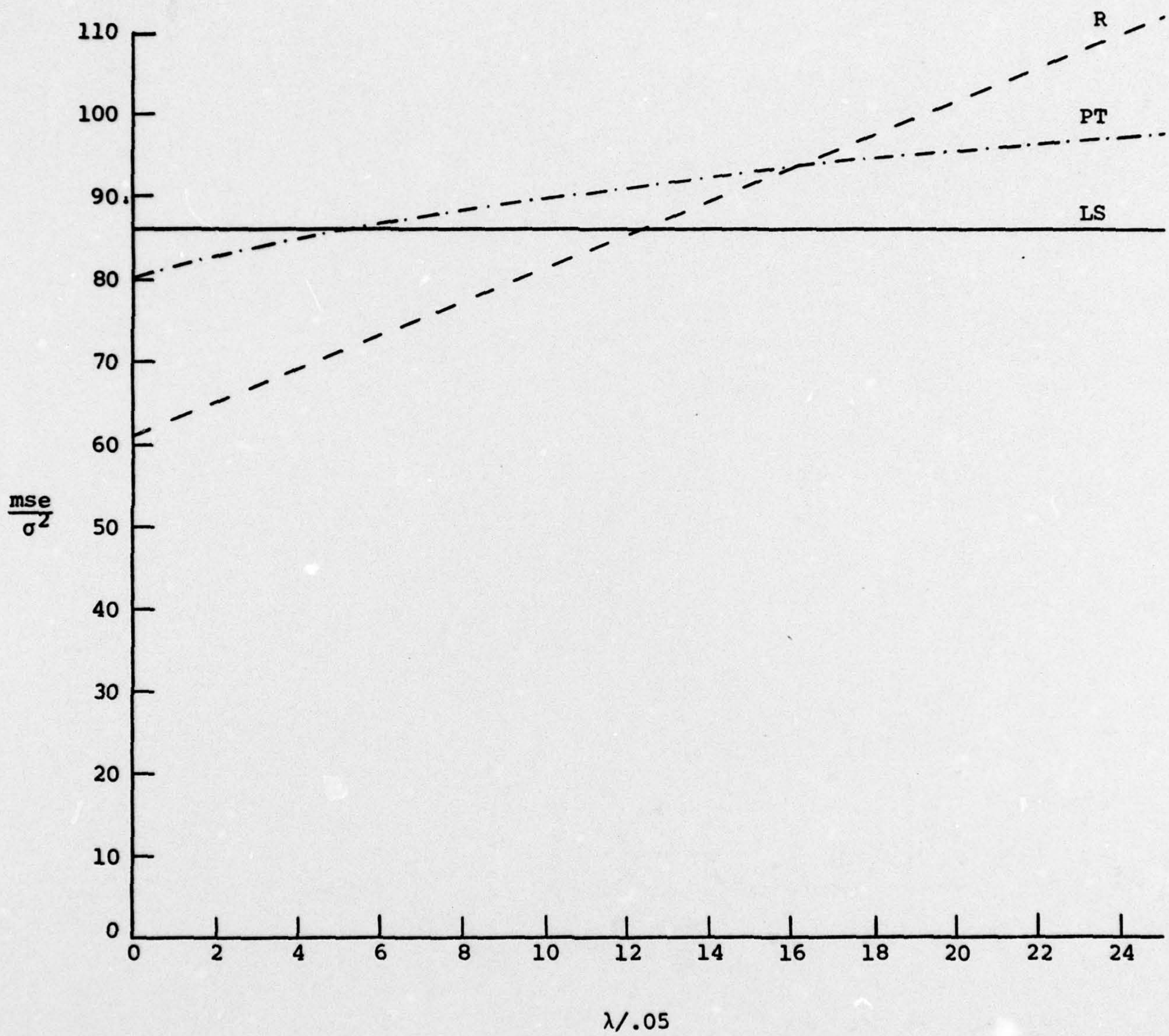
ℓp=1, α=.25
ℓp=1, α=.05
ℓp=.01, α=.25
ℓp=.01, α=.05
Power
1.0
.9
.8
.7
.6
.5
.4
.3
.2
.1
0
0 1 2 3 4 5 6 7 8 9 10
λ/ℓp



mse
120
110
100
90
80
70
60
50
40
30
20
10
0
0
1
2
3
4
5
6
7
8
9
10
λ/ℓp
LS, ℓp=.01
PT, ℓp=.01
R, ℓp=.01
R, ℓp=1
PT, ℓp=1
LS, ℓp=1



*Both curves for R are so close only one is plotted.


R
PT
LS
110
100
90
80
70
60
50
40
30
20
10
0
mse
σ²
0
2
4
6
8
10
12
14
16
18
20
22
24
λ/.05

REFERENCES

Bock, M. E., Yancey, T. A., and Judge, G. G. (1973). "Statistical Consequences of Preliminary Test Estimators in Regression," Journal of the American Statistical Association, 68, 109-16.

Dempster, A. P., Schatzoff, M., and Wermuth, N. (1977). "A Simulation Study of Alternatives to Ordinary Least Squares," Journal of the American Statistical Association, 72, 77-90.

Gunst, R. F. and Mason, R. L. (1977). "Biased Estimation in Regression: An Evaluation Using Mean Squared Error," Journal of the American Statistical Association, 72, 616-28.

Jeffers, J. N. R. (1967). "Two Case Studies in the Application of Principal Component Analysis," Applied Statistics, 16, 225-36.

Lawless, J. F. and Wang, P. (1976). "A Simulation of Ridge and Other Regression Estimators," Communications in Statistics, A5, 307-23.

Lott, W. F. (1973). "Optimal Set of Principal Component Restrictions on a Least Squares Regression," Communications in Statistics, 2, 449-64.

Mansfield, E. R. (1975). "Principal Component Approach to Handling Multicollinearity in Regression Analysis," Ph.D. Dissertation, Department of Statistics, Southern Methodist University, Dallas, Texas.

Mason, R. L., Gunst, R. F., and Webster, J. T. (1975). "Regression Analysis and Problems of Multicollinearity," Communications in Statistics, 4, 277-92.

Massy, W. F. (1965). "Principal Component Regression in Exploratory Statistical Research," Journal of the American Statistical Association, 60, 234-56

Silvey, S. P. (1969). "Multicollinearity and Impercise Estimation," Journal of the Royal Statistical Society, Series B, 31, 539-52.

Toro - Vizcarrondo, C. and Wallace, T. D. (1968). "A Test of the Mean Square Error Criterion for Restrictions in Linear Regression," Journal of the American Statistical Association, 63, 558-72.